AN AUTOMATED DIFFERENTIAL BLOOD COUNT SYSTEM

Guclu Ongun[(1)], Ugur Halici[(1)], Kemal Leblebicioglu[(1)], Volkan Atalay[(2)],
Meral Beksac[(3)], Sinan Beksac[(4)]

(1) Computer Vision and Neural Networks Research Center,
Dept. of Electrics and Electronics Eng., Middle East Technical University, 06531, Ankara, Turkey
guclu.ongun@isbank.net.tr, halici@metu.edu.tr, kleb@metu.edu.tr
(2) Dept. of Computer Eng., Middle East Technical University, 06531, Ankara, Turkey
(3) Faculty of Medicine, Ankara University, Ankara, Turkey
(4) Faculty of Medicine, Hacettepe University, Ankara, Turkey

Abstract **-While the early diagnosis of hematopoietic system disorders is very important in hematolgy, it is a highly complex and time consuming task. The early diagnosis requires a lot of patients to be followed-up by experts which, in general is in-feasible because of the required number of experts. The differential blood counter (DBC) system that we have developed is an attempt to automate the task performed manually by experts in routine. In our system, the cells are segmented using active contour models (snakes and ballons), which are initialized using morphological operators. Shape based and texture based features are utilized for the classification task. Different classifiers such as k-nearest neighbors, learning vector quantization, multi-layer perceptron and support vector machine are employed.**

Keywords **Differential blood counter, cell recognition, active contours, snakes, neural networks, support vector machine**

I. INTRODUCTION

An important issue in hematology is the early diagnosis of hematopoietic system disorders (HSD). Since HSD are critical, examination requires expert evaluation and is a highly complex and time consuming task. White cell composition of the blood reveals important diagnosis information about the patients as well as patient follow-up. The hematologist requires two types of blood count for diagnosis and screening. The first one is called the Complete Blood Count (CBC) and the second one is called the Differential Blood Count (DBC). CBC could be done by instruments called cytometer and could successfully be performed automatically. On the other hand, DBC is more reliable but currently it is a manual procedure to be done by hematology experts using microscope. In DBC, an expert counts 100 white blood cells on the smear at hand and computes the percentage of occurrence of each type of cell counted. The results reveal important information about patient's health status. Apparently, DBC is a time consuming task that requires expert examination.

Our automated differential blood counter system is an attempt for performing DBC automatically by the aid of statistical and neural network based classification methods.

The process of counting blood cells on smear images requires four steps. These steps are acquisition, segmentation, feature extraction, and classification.

Very few methods are presented in the literature for the segmentation step. Morphological preprocessing followed by fuzzy-patch labeling is proposed in [1] for locating the white blood cells. Then, the nucleus centers are detected by variance map and it is followed by a snake-based segmentation. In [2], we had used contour following to segment the cell groups and then used the curvature to seperate the overlapping cells. In [3], we combined snakes with balloons for segmenting cells directly. In other related papers, segmentation is done manually.

In feature extraction step, intensity-based features are used in common [4-6]. However, some authors prefer to use texture-based features, and/or shape descriptors [4, 5].

For the classification, neural network based classifiers are used in [2,5,6]. Due to the fuzzy nature of the decision process in counting blood cells, a dedicated neural network counter is constructed in [5]. In this work, the authors state the fact that the results of a counting session could be different between trials about 15%.

In order to conduct an automated counter, methods performing well for segmentation, feature extraction, and classification are needed. In our current system, segmentation is done by morphological preprocessing followed by the snake-balloon algorithm.

Several types of features such as intensity and color based features, texture based features, and shape based features are utilized for a robust representation of the objects.

For classification we employed k-Nearest Neighbors (k-NN), Learning Vector Quantization (LVQ), Multi-Layer Perceptron (MLP) and Support Vector Machine (SVM).

The organization of the paper is as follows: In section 2, the blood cell image database that we collected and the cell categories that we considered are explained. In section 3, the

Report Documentation Page

Report Date	Report Type	Dates Covered (from... to)
25 Oct 2001	N/A	-

Title and Subtitle An Automated Differential Blood Count System	**Contract Number**
	Grant Number
	Program Element Number
Author(s)	**Project Number**
	Task Number
	Work Unit Number
Performing Organization Name(s) and Address(es) Computer Vision and Neural Networks Research Center Dept of Electrics Eng Middle East Technical University 06531, Ankara, Turkey	**Performing Organization Report Number**
Sponsoring/Monitoring Agency Name(s) and Address(es) US Army Research, Development & Standardization Group (UK) PSC 803 Box 15 FPO AE 09499-1500	**Sponsor/Monitor's Acronym(s)**
	Sponsor/Monitor's Report Number(s)

Distribution/Availability Statement
Approved for public release, distribution unlimited

Supplementary Notes
Papers from 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, October 25-28, 2001, held in Istanbul, Turkey. See also ADM001351 for entire conference on cd-rom., The original document contains color images.

Abstract

Subject Terms

Report Classification unclassified	**Classification of this page** unclassified
Classification of Abstract unclassified	**Limitation of Abstract** UU

Number of Pages
4

architecture of the system is given. In section 4, the segmentation procedure is presented. Section 5 focuses on feature extraction and types of features used in the process. Then in section 6, the classification results for different methods are explored. The last section concludes the study.

II. BLOOD CELL IMAGE DATABASE

All the cell classes are evolved from a single young cell produced in bone marrow due to different bio-chemical reactions. In that sense, cell classes form a family tree.

The following cell classes are important in terms of DBC. In bone marrow: Erythroblast, Lymphoblast, Metamyelocyte, Monoblast, Myeloblast, Myelocyte, Plasma cell, Proerthroblast, Promyelocyte, Band, and Megakaryocyte. In peripheral blood: Neutrophil, Basophile, Eisonophil, Lymphocyte, and Monocyte. It should be noted that eritrocytes, which appear in peripheral blood, have no importance on DBC.

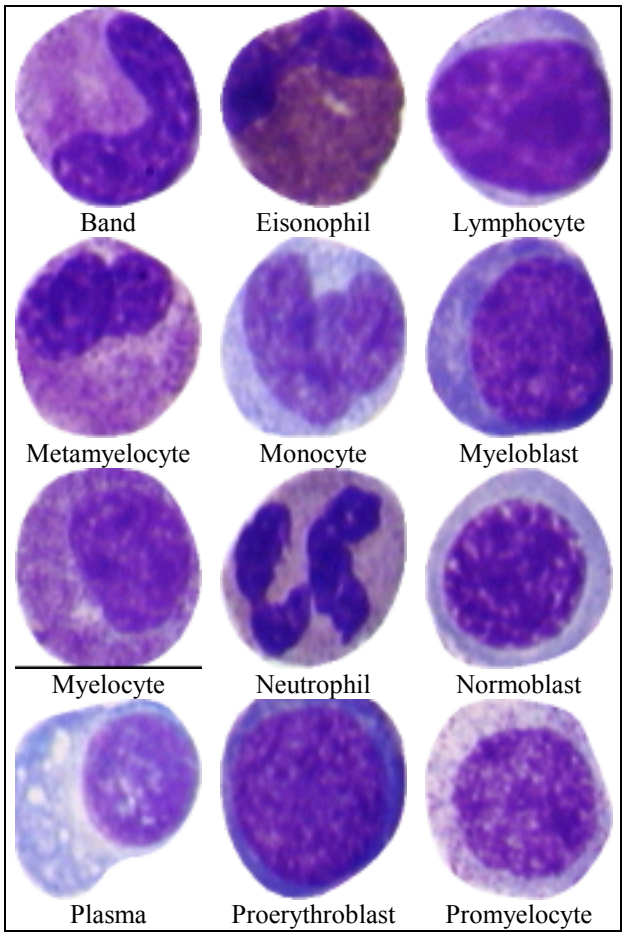


Fig 1. Samples of white blood cells

There are two mediums in which the white cells can be analyzed. Bone marrow is the production and maturing place for the cells. After the cells reach certain maturity level, they are released to blood to perform certain tasks. Detection of immature cells in peripheral blood signals a problem in an individual's health status [7].

Our blood cell image database has been constructed at the hematology laboratory of Hacettepe University Hospital, Ankara. The database contains 108 images of 258 white cells, most of them being bone marrow images and these cells are classified manually.

Not all the sixteen classes listed above, but twelve of them (given in Figure 1) are considered in this study. The other four classes are not taken into account due to insufficient number of samples in the database. It should be noted that, the cell images given in Figure 1 are scaled to have approximately the same width and height for display purposes. Actually, this is not the case in the microscopic images and cell area is a component in feature vector, as it will be explained in section 5.

III. SYSTEM ARCHITECTURE

The architecture of our system is as follows: As the input device supplies an image, cell segmentation procedure is carried out. Segmentation yields to a number of cell contours and their nuclei. Then, the feature extraction engine analyzes each segmented cell and its nucleus to form a feature vector from color, shape, and texture features. Feature vectors are stored to constitute the dataset. Training and testing sets are chosen to be mutually exclusive. Classifiers are constructed by using the training set as input to the given classification methods. After a classifier is constructed, test images are analyzed and each object in these images are labeled by the classifier according to their feature vectors.

IV. SEGMENTATION

In order to locate the cells for feature extraction, we have used active contour models, widely known as snakes. Method of snakes is successfully used in detecting contours of the objects in multi-valued images (i.e. grayscale, color, volume data, etc) [1], [8-16].

An active contour is an energy-minimizing curve defined as follows:

$$E^*_{snake} = \int_0^1 E_{\text{int}}(v(t)) + E_{ext}(v(t))dt$$

$$= \int_0^1 \left[E_{\text{int}}(v(t)) + E_{image}(v(t)) + E_{other}(v(t))\right]dt$$

where

$$E_{image} = w_{line}E_{line} + w_{edge}E_{edge} + w_{term}E_{term}$$

$$E_{line} = I(x,y)$$

$$E_{edge} = -\left|\nabla I(x,y)\right|^2$$

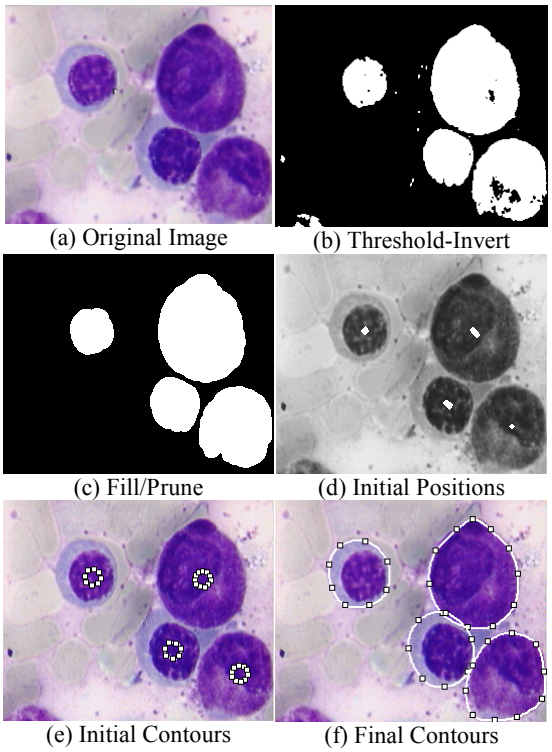
$$E_{term} = \frac{\partial v}{\partial n_R} = \frac{\partial^2 v / \partial n_R^2}{\partial v / \partial n}$$

Behavior of the snake is controlled by adjusting w_{line}, w_{edge}, and w_{term}. The termination energy is not used in this work. E_{other} could be used for application specific purposes, e.g., it helps snake to be able to inflate in the case of balloons [9].

One important drawback of the original snake algorithm is initial positioning. Several methods are proposed for minimizing the effect of initial positioning, such as segmented snakes [15], dual active contours [12] and gradient vector flow snakes [11]. Convexity analysis of energy minimization is done in [16].

In this work, initial positioning is performed by making use of digital morphology. The algorithm for finding initial position of the snake is outlined below:

1. Convert the color image to intensity image;
2. Sub-sample intensity image;
3. Find a threshold that yields a mask containing cell nucleuses;
4. Perform closing to smooth out the mask;
5. Perform distance transform and find relative maxima on the mask;
6. Label and merge the maxima regions;
7. Compute the center of mass of merged regions that yields the initial position for the snake.



(a) Original Image (b) Threshold-Invert
(c) Fill/Prune (d) Initial Positions
(e) Initial Contours (f) Final Contours

Fig. 2. Segmentation steps.

After the initial positions are found, snakes are put on the image and minimization procedure is performed. Upon convergence, the interior of each contour is taken as a white blood cell. For finding nucleus region(s), the constraints of energy functional are changed. Initial snakes are chosen as the contours of the cells found in the previous step. This procedure is demonstrated in Figure 2. The details of this fast snake-balloon method that we have developed for segmentation of cells can be found in [3].

It should be noted that, in the segmentation procedure, only white blood cells are segmented and the other objects, including erythrocytes, are eliminated.

V. Feature Extraction

Since the chosen features affect the classifier performance much, deciding on which features to be used in a specific data classification problem is as important as the classifier itself. In this work, we tried to reflect the rules and heuristics used by the hematology experts in the selected features.

Our features mainly fall into two categories: shape based features and color/texture based features.

For classifying cells successfully, hematology experts examine the shape of the cells and nuclei. To reflect this information in our feature vectors, several tools such as moments and affine invariants are taken from the literature [17,18] together with some additional features that are heuristically picked by analyzing the reasoning of hematology experts. These additional features include the areas of cell and nucleus; ratios of nucleus area and perimeter length over cell area and perimeter length, respectively; compactness and boundary energy of nucleus; nucleus shape.

As color and texture features, mean and standard deviation for cell, cytoplasm, and nucleus in CIE (L^*,a^*,b^*) color system and also histograms in Hue-Saturation-Value (HSV) color system are used [19].

Totally we have used a 57 dimensional feature vector. The details related to these features are presented in [4]

VI. Classification

The classification algorithms that we tested on our blood cell image database are k-Nearest Neighbor (k-NN) [19], Linear Vector Quantization (LVQ) [21], Multi Layer Perceptron (MLP) [20] and Support Vector Machines Machine (SVM) [22-26]

The classification accuracy of the methods mentioned above are computed by selecting random non-intersecting training and test sets, such that the training set consist of 70% of the dataset, the number of samples from each class being

proportional to their number in the whole dataset, and the remaining %30 of the dataset is taken as the test set. For each method, the experiment is repeated with 100 random training and test sets. The best performances of the methods in these 100 experiments are as follows:
Training accuracy is 82%, 94%, 99% and 100% for k-NN, LVQ, MLP and SVM, respectively. The corresponding performances on test sets are 81%, 83%, 90% and 91%.

VII. CONCLUSIONS

In this paper, the segmentation, feature extraction and classification phases for the automated DBC system that we developed is presented. The performance of the system is encouraging. Currently, we are working on evaluating classifier combinations such as committees of networks [20] and stacked generalization [27] to improve the robustness of the classification step.

REFERENCES

[1] J. Park and J. Keller, "Fuzzy Patch Label Relaxation in Bone Marrow Cell Segmentation", *IEEE International Conference on Systems, Man, and Cybernetics*, pp. 1133-1138, 1997.

[2] G. Ongun, U. Halici, K. Leblebicioglu, V. Atalay, A. Erkmen, S. Beksaç, M. Beksaç, A. Erol, "An Automated Blood Cell Analysis and Classification System", *Proc. 13th International Symposium on Computer and Information Sciences*, IOS Press, Antalya , Turkey, Oct. 1998, pp 222-227

[3] G. Ongun, U. Halici, K. Leblebicioğlu, V. Atalay, S. Beksac, and M. Beksac, "Automated Contour Detection in Blood Cell Images by An Efficient Snake Algorithm", *Nonlinear Analysis: Series B - Applications*, (in press).

[4] Ongun G., Halici U., Leblebicioglu K., Atalay V., Beksaç M., Beksaç S., "Feature Extraction and Classification of Blood Cells for an Automated Differential Blood Count System, *Proc. of IJCNN 2001*, Wahington DC, USA (in press).

[5] N. Theera-Umpon and P.D. Gader, "Training Neural Networks to Count White Blood Cells via a Minimum Counting Error Objective Function", *Proc. Int'l Conf. on Pattern Recognition*, 2000.

[6] M. Beksac, M. S. Beksac, V.B. Tipi, H.A. Duru, M.U. Karakas, and A. Cakar, "An Artificial Intelligent Diagnostic System on Differential Recognition of Hematopoietic Cells From Microscopic Images", *Cytometry 30*, pp. 145-150, 1997.

[7] A.V. Hoffbrand and J.E. Pettit, *Clinical Heamatology*, Times Mirror International Publisher Ltd., second ed., 1994.

[8] M. Kass, A. Witkin, and D. Terzopoulos, Snakes: Active Contour Models, *International Journal of Computer Vision*, No. 4, pp 312-331, 1988.

[9] L. D. Cohen, I. Cohen, "Finite-Element Methods for Active Contour Models and Balloons for 2-D and 3-D Images", *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 15, No. 11, pp. 1131-1147, 1993.

[10] L.D. Cohen and R. Kimmel, "Global Minimum for Active Contour Models: A Minimal Path Approach", *IEEE Comp. Soc. Conf. on Computer Vision and Pattern Recognition*, pp. 666-673, 1996.

[11] C. Xu, J. L. Prince, "Snakes, Shapes, and Gradient Vector Flow", *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 7, No. 3, pp.359-369, 1998.

[12] S.R. Gunn and M.S. Nixon, "A Robust Snake Implementation: A Dual Active Contour", *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 1, pp. 63-69, 1997.

[13] B. Olstad and A.H. Torp, "Encoding of a priori Information in Active Contour Models", *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. 18, No.9, pp. 863-872, 1996.

[14] G. Sapiro, "Vector-Valued Active Contours", *IEEE Comp. Soc. Conf. on Computer Vision and Pattern Recognition*, pp. 680-685, 1996.

[15] Y.Y. Wong, et al., "Segmented Snake for Contour Detection", *Pattern Recognition*, Vol. 31, No. 11, pp. 1669-1679, 1998.

[16] C. Davatzikos and J.L. Prince, "Convexity Analysis of Active Contour Problems", *IEEE Comp. Soc. Conf. on Computer Vision and Pattern Recognition*, pp. 674-679, 1996.

[17] A.K. Jain, *Fundementals of Digital Image Processing*, Prentice Hall, E. Clifs, NJ, 1989.

[18] J. Flusser and T. Suk, "Pattern Recognition by Affine Moment Invariants", *Pattern Recognition*, Vol. 26, No. 1, pp. 167-174, 1993.

[19] M. Unser, "Sum and Difference Histograms for Texture Classification", *IEEE Transactions on Pattern Recognition and Machine Intelligence*, Vol. 8, pp. 118-125, 1986.

[20] T. Kohonen, *Self-Organizing Maps*, Springer Series in Information Sciences, Berlin, Vol. 30, 1995.

[21] C.M. Bishop, *Neural Networks for Pattern Recognition*, Clarendon Press, Oxford, 1995.

[22] C.J.C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition", *Data Mining and Knowledge Discovery*, Vol. 2, No. 2, pp. 121-167, 1998.

[23] M.A. Hearst, "Support Vector Machines", *IEEE Intelligent Systems*, pp. 18-28, 1998.

[24] E. Osuna, R. Freund, and F. Girosi, "An Improved Training Algorithm for Support Vector Machines", *Proc. IEEE Workshop on Neural Networks for Signal Processing,* 1997.

[25] J. Platt, *Fast Training of Support Vector Machines Using Sequential Minimal Optimization*, MIT Press, Cambridge, Mass., Advances in Kernel Methods: Support Vector Learning, 1999.

[26] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer, 1995.

[27] D.H. Wolpert, "Stacked Generalization", *Neural Networks,* Vol. 2, pp. 241-259, 1992.